# Metacognitive Continual Learning: A Research Agenda for Post-Scaling Training Architectures

Declan Michaels | moral-os.com | June 2026

## The Problem

Scaling is approaching its practical limits. Data, compute, money, power, politics. The path that got us here is not the path that gets us further. We need to think differently about how we use the resources available to us.

Models spend enormous compute during inference and routinely construct novel representations and reasoning pathways in the process. Within a session, a model builds context, discovers connections, arrives at formulations it could not have produced at session start. All of it is wiped at session end. At the scale of millions of daily sessions, this is a substantial missed opportunity.

## The Claim

Ordered curriculum builds more structured representational geometry. Models trained on identical data in different orders develop measurably different internal structure. Those differences survive post-hoc alignment. Behavioral fine-tuning does not restructure them. (A scaled replication is in progress.)

If that finding holds at scale, the implication is broader than curriculum design. It means that pretraining, fine-tuning, and continual learning are all instances of the same underlying process: ordered geometric development across time. Pretraining controls which structures form first. Fine-tuning and RLHF reshape behavior without restructuring what is underneath. What has been missing is a mechanism for structural modification after deployment, one that operates on geometry, not just outputs.

This brief proposes that mechanism.

## The Architecture

A closed-loop architecture with three components:

**Retrospective Flagging.** At session end, the model asks: what broke, what worked, what's new.

**Mentor-Guided Diagnosis.** A separate model reviews flagged transcripts and engages the target model in a diagnostic session, with access to internal activations during replay. The mentor works at the representation level: a failure might reveal a missing connection, a success might reveal an emergent structure worth reinforcing before it is lost.

**Targeted Consolidation.** Flagged experiences enter a shared queue. The mentor evaluates the queue in categorized batches, filtering for patterns that recur across sessions. Patterns that pass are replayed during a dedicated consolidation phase with elevated learning rates. Because a mentor targets and supervises every change, consolidation can run opportunistically and deploy incrementally, when electricity is cheap and demand is low.

The result is a model that participates in directing its own learning. An external mentor guides the process, but the signal originates from the model's own experience. The fixed curriculum of pretraining gives way to a continuous one, shaped by what the model encounters and what the mentor diagnoses.

## Experimental Approaches

Three approaches to the weight modification step, in increasing order of technical difficulty and potential impact. They are not alternatives. They are points on a continuum from indirect to direct geometric intervention.

**Approach 1: Targeted Fine-Tuning.** The mentor generates training examples that exercise weaknesses or reinforce emergent strengths. Standard supervised fine-tuning handles the weight update. Geometric change is a side effect of example selection.

**Approach 2: Activation-Informed Curriculum.** During replay, the mentor can see where in the network structure broke down or formed. It then designs example sequences that progressively build or strengthen that structure, verifying at the representation level whether each intervention is working. Geometric change is the explicit objective, pursued indirectly through curriculum.

**Approach 3: Direct Geometric Intervention.** The mentor identifies specific geometric features and constructs loss objectives targeting them directly: strengthening cross-domain connections, stabilizing emergent clusters, restructuring layer geometry. Geometric change is both the objective and the method. The most technically demanding approach, likely requiring tooling and techniques that do not yet exist.

## Open Questions

**Flagging quality.** How noisy is self-assessment? The mentor and recurrence filtering compensate for false positives, but the efficiency of the pipeline depends on the model's ability to flag signal over noise. Whether that ability improves through the consolidation loop is an empirical question.

**Mentor architecture.** Is a peer model with different training sufficient, or must the mentor be more capable than the target? A peer of a different architecture would reduce shared blind spots and self-reinforcing biases. If a peer can serve, the architecture becomes accessible without requiring a larger model.

**Intentional forgetting.** When a human recognizes their approach as flawed and sets out to learn a better one, the forgetting is targeted and intentional. Can consolidation support deliberate unlearning alongside reinforcement?

**Privacy.** The queue contains flagged user sessions. All standard privacy requirements apply: storage, anonymization, and consent. These are not research questions, but they constrain every design decision upstream.

**Scale thresholds.** At what model size does metacognitive flagging produce reliable signal?

**Retrieval augmentation.** The mentor needs context during diagnosis. Retrieval-augmented generation could give the mentor access to prior flagged sessions, known failure patterns, or external knowledge without requiring all of it to live in weights. How retrieval integrates with activation-level diagnosis is an open design question. It may also change what needs to be consolidated at all: if retrieval can compensate for a gap, the consolidation queue shrinks.

## Conclusion

Models routinely construct novel reasoning pathways during inference that vanish with the session. This architecture proposes the mechanism to keep them: a closed loop from self-assessment through structural diagnosis to targeted consolidation.

The deeper claim is that this is not a new kind of learning. It is the same process that operates during pretraining: ordered geometric development across time. The difference is who directs it. In pretraining, the curriculum is fixed externally. In the architecture proposed here, the model contributes to its own curriculum through experienced failures, emergent successes, and mentor-guided diagnosis. The learning process is the same. The locus of control shifts.

Cite as: Michaels, D. (2026). Metacognitive Continual Learning: A Research Agenda for Post-Scaling Training Architectures. Zenodo. https://doi.org/10.5281/zenodo.20799680 · 